

基于人工智能的海上油田生产系统故障预测模型研究

田国栋 张 阳 吕寻凯 王立新
中海油能源发展装备技术有限公司 天津 300452

摘要：随着海洋油气资源开发力度不断加大，海上油田生产系统的安全稳定运行成为行业关注焦点。本文聚焦于基于人工智能的海上油田生产系统故障预测模型研究。首先概述海上油田生产系统，接着阐述人工智能技术在故障预测领域的应用。随后详细介绍基于人工智能的故障预测模型构建过程，包括数据采集与预处理、特征工程、模型选择与训练以及模型评估与优化。旨在通过人工智能技术提升海上油田生产系统故障预测的准确性与及时性，保障生产安全与稳定运行。

关键词：基于人工智能；海上油田生产；系统故障预测；模型研究

引言：海上油田作为国家重要的能源产出地生产基地，其生产系统的稳定运行至关重要。然而，由于海上环境复杂，设备长期处于恶劣工况，故障频发，不仅影响生产效率，还可能引发安全事故，造成巨大经济损失。传统的故障预测方法多依赖人工经验与定期检修，存在预测精度低、时效性差等问题。随着人工智能技术的飞速发展，其在故障预测领域展现出巨大潜力。通过利用人工智能算法对海量生产数据进行分析挖掘，能够更精准地预测故障发生，提前采取维护措施。因此，开展基于人工智能的海上油田生产系统故障预测模型研究具有重要的现实意义与应用价值。

1 海上油田生产系统概述

海上油田生产系统是一个集成多种设备与技术的复杂工程体系，旨在实现海洋油气资源的开采、处理、储存与运输，其结构与运行模式需适应严苛的海洋环境与高效生产需求。该系统主要由井口开采系统、集输处理系统、储油与外输系统及辅助支持系统四部分构成：井口开采系统通过水下井口、采油树等设备将油气从地层输送至平台；集输处理系统对油气进行分离、净化与计量，去除水分、杂质及有害成分；储油与外输系统通过储油舱或浮式储油装置暂存原油，并借助管线或穿梭油轮完成外输；辅助支持系统则涵盖动力供应、通信导航、安全监控等设施，保障主系统稳定运行。与陆地油田相比，海上生产系统具有显著特殊性：一方面，其长期暴露于高盐雾、强风浪、深水压等极端环境中，设备腐蚀、疲劳损伤速率远高于陆地；另一方面，系统设备密集且空间受限，各子系统间耦合度高，单一设备故障易引发连锁反应。此外，海上作业成本高昂、维修难度大，一旦发生故障，不仅可能导致生产中断，还可能引发漏油等环境灾难。传统运维模式下，依赖定期检修与

人工巡检，难以实时捕捉设备隐性故障，且应对突发状况的响应效率较低。因此，深入理解该系统的结构特性与运行挑战，是构建精准故障预测模型的基础，也是推动其向智能化运维转型的前提^[1]。

2 人工智能技术在故障预测中的应用

2.1 机器学习算法

机器学习算法是一类通过分析历史数据构建预测模型的数据方法，在海上油田故障预测领域有着切实的应用价值。这类方法能够从设备运行参数、故障记录等结构化数据中，自主挖掘潜在的规律和关联，从而实现对故障的提前判断。常见的包括基于统计分析的支持向量机、基于决策规则的决策树，以及集成多种模型优势的随机森林等。在海上油田生产系统中，机器学习算法可处理井口压力、泵体温度等多维度的监测数据。其特点是计算过程清晰明了，得出的结果易于理解和解释，能够适应海上平台有限的运算资源。不过，在处理高维度且具有非线性特征的数据时，其精准度会受到一定程度的限制，这是该类算法在实际应用中需要注意的地方。

2.2 深度学习算法

深度学习算法依托多层网络结构实现对复杂数据的解析，能够有效处理海上油田生产过程中的非结构化信息和高维度参数。它通过模拟数据处理的多层次转换过程，能够自动提取数据中隐藏的特征，无需人工预先设定分析规则。典型的如擅长处理图像数据的卷积网络，可识别设备表面相关的视觉特征；适用于时序数据的循环网络，能捕捉压力、流量等参数随时间变化的异常模式。在海上油田场景中，这类算法可整合传感器实时数据、设备运行日志等多源信息。其优势在于能突破人工特征筛选的局限，提升对复杂故障的识别能力。但它也

存在一定不足，需要大量标注数据作为支撑，且运算过程较为复杂，得出结果的直观解释性相对较弱。

2.3 知识图谱技术

知识图谱技术是将设备信息、故障案例、专家经验等知识进行结构化整合的方法，通过构建实体与关系的关联网络，实现对故障的推理分析。它把设备型号、故障类型、成因特征等要素转化为可计算的关联关系，形成一个完整的知识体系。在海上油田生产系统中，知识图谱可整合设备手册中的参数标准、运维记录中的故障处理经验。这种方法弥补了单纯依赖数据的预测方式对隐性知识的忽视，尤其适用于多因素交织的复杂故障分析。同时，它能让预测结果的逻辑更加清晰，便于现场工作人员理解和应用，为故障预测领域提供了一种坚实的知识支撑方法。

2.4 强化学习算法

强化学习算法通过动态调整决策策略来实现预测精度的优化，其核心在于在与运行环境的持续交互过程中，根据反馈不断修正分析模型，使预测结果更贴合实际的工况。它通过设定目标导向的评价机制，引导模型在设备运行状态与故障判断标准之间建立动态的适配关系。在海上油田生产中，这类算法能够应对海洋环境变化给设备运行带来的波动。同时，它能结合故障预测结果对维护方案进行优化，提高海上平台的应急处理效率。不过，其应用需要积累大量的实际运行数据，而且在高风险场景中，需要先通过模拟测试验证效果，之后再逐步应用于实际生产过程中^[2]。

3 基于人工智能的海上油田生产系统故障预测模型构建

3.1 数据采集与预处理

3.1.1 数据采集

数据采集是故障预测模型构建的基础，需覆盖海上油田生产系统的全流程运行信息。采集对象包括井口开采系统的压力、温度、流量等实时参数，通过部署在水下井口、采油树等设备上的传感器实现高频次监测；集输处理系统的分离设备运行数据、介质成分数据，依托在线分析仪器完成自动记录；储油与外输系统的液位、压力、阀门状态等信息，通过控制系统实时抓取；辅助系统的动力供应参数、环境监测数据（如风速、波浪高度）也需同步采集。此外，还需整合设备历史故障记录、维护日志、出厂参数等静态数据，形成多维度数据集。

3.1.2 数据清洗

数据清洗旨在去除原始数据中的异常值、缺失值与噪声，为模型提供可靠输入。针对海上油田生产数据的

特点，采用统计分析方法识别异常值，如通过 3σ 原则筛选超出合理范围的压力、温度数据，结合设备运行逻辑判断突变值是否为真实故障信号；对于缺失值，根据数据特性选择插值法（如线性插值、多项式插值）或基于相邻时段数据的填补方法，避免因数据断裂影响模型训练；针对传感器漂移、电磁干扰产生的噪声，采用滑动平均、小波变换等滤波技术平滑数据曲线。

3.1.3 数据标准化

数据标准化通过统一不同类型数据的量纲与分布范围，提升模型对特征的识别效率。海上油田生产数据中，压力、温度、流量等参数的单位与数值范围差异较大，需通过标准化处理消除量纲影响。常用方法包括Min-Max标准化，将数据映射至 $[0,1]$ 区间，适用于分布较为均匀的参数；Z-Score标准化，将数据转换为均值为0、标准差为1的正态分布，适合存在极端值的数据集。对于分类数据（如设备状态标签、故障类型），采用独热编码或标签编码转换为数值形式。标准化过程需基于训练数据集的统计特征制定转换规则，确保测试数据与训练数据采用一致的标准化标准，避免因数据分布差异影响模型泛化能力。

3.2 特征工程

3.2.1 特征提取

特征提取是从原始数据中挖掘与故障相关的关键信息，将高维或非结构化数据转化为可用于模型训练的低维特征。针对海上油田生产系统的多源数据，需结合数据类型采用差异化方法：对传感器采集的时序数据（如压力波动、振动信号），通过时域分析提取均值、方差、峰值等统计特征，或利用傅里叶变换、小波变换转换至频域，捕捉设备异常的频率特征；对设备运行日志等文本数据，通过关键词提取与语义分析，转化为反映故障关联的特征向量；对图像数据（如设备外观检测图），采用边缘检测、纹理分析等方法提取形态特征。特征提取需聚焦故障敏感性指标，例如通过提取泵体振动的高频分量识别轴承早期磨损，通过管道压力的突变特征捕捉泄漏风险，为后续模型提供具有物理意义的输入变量。

3.2.2 特征选择

特征选择旨在从提取的特征集中筛选出最具代表性的子集，减少冗余信息对模型的干扰，提升训练效率与预测精度。海上油田生产系统的特征选择需结合领域知识与算法分析：基于专家经验筛选与设备故障直接相关的特征，如井口压力、温度变化率等核心参数；采用统计方法（如方差分析、互信息）衡量特征与故障标签

的相关性，剔除低关联度特征；通过机器学习算法（如随机森林、递归特征消除）评估特征对模型性能的贡献度，保留重要特征。选择过程需平衡特征数量与模型复杂度，避免因特征过少导致信息缺失，或因特征过多引发过拟合。

3.3 模型选择与训练

3.3.1 模型选择

模型选择需结合海上油田生产系统的故障特征与数据特性，兼顾预测精度与工程实用性。针对系统中设备的早期微弱故障，可选用随机森林、梯度提升树等集成学习模型，其对非线性关系的捕捉能力较强，且能处理混合类型特征，适合井口压力、振动信号等多源数据的融合分析；对于具有明显时序特性的故障（如管道腐蚀的渐进式变化），长短期记忆网络更能捕捉参数随时间的动态关联，提升趋势预测的准确性；若需兼顾模型解释性以辅助现场决策，可优先选择决策树或逻辑回归等白盒模型，便于工程师理解故障判断依据。同时，需考虑海上平台的算力限制，避免选择结构过于复杂的模型，通过对比不同模型在验证集上的表现（如准确率、召回率），确定适配特定故障类型的最优模型。

3.3.2 模型训练

模型训练是通过优化参数使模型拟合故障数据规律的过程，需以标准化后的特征数据为输入，构建“特征-故障标签”的映射关系。训练前需将数据集按比例划分为训练集（70%-80%）与验证集（20%-30%），前者用于模型参数学习，后者用于监控训练过程中的过拟合风险。训练过程中，针对分类任务（如故障有无）采用交叉熵损失函数，针对回归任务（如故障发生时间）采用均方误差损失函数，通过梯度下降算法迭代优化参数。需重点控制训练轮次，当验证集性能不再提升时及时停止，避免模型过度拟合训练数据的噪声。同时，引入早停机制与正则化方法，增强模型对新数据的适应能力，确保训练后的模型能稳定应用于海上油田复杂多变的实际工况。

3.4 模型评估与优化

3.4.1 模型评估

模型评估需通过多维度指标全面衡量预测性能，结合海上油田生产系统的故障风险特性选择适配标准。针对故障分类任务，采用准确率反映整体判断正确性，同时重点关注召回率（避免漏报故障）与精确率（减少误

报干扰），尤其对可能引发安全事故的关键故障（如井口泄漏），需将召回率作为核心指标。对于故障发生时间的回归预测，采用均方根误差、平均绝对误差评估数值偏差，结合实际故障间隔计算预测提前量，确保模型能预留足够的应急处理时间。评估过程需基于独立测试集展开，通过与历史故障记录对比，验证模型在不同工况（如极端天气、设备高负荷运行）下的稳定性，同时引入混淆矩阵、ROC曲线等可视化工具，直观呈现模型的优缺点，为后续优化提供明确方向。

3.4.2 模型优化

模型优化旨在通过迭代改进提升预测性能，需针对评估中暴露的问题采取针对性措施。若模型泛化能力不足，可通过增加训练数据量（尤其是罕见故障样本）、采用数据增强技术扩充数据集，或引入迁移学习方法利用相似设备的故障数据辅助训练。针对过拟合问题，可调整模型复杂度（如简化神经网络层数、增加正则化系数），或采用集成学习融合多个弱模型的预测结果，平衡偏差与方差。若关键指标（如召回率）不达标，可通过调整损失函数权重强化对高风险故障的关注，或优化特征工程环节，补充与故障强相关的新特征。优化过程需结合现场反馈，定期将实际运行数据输入模型进行再训练，使模型适应设备老化、工况变化等动态因素，维持长期稳定的预测效果^[3]。

结束语

本研究围绕基于人工智能的海上油田生产系统故障预测模型展开，通过整合机器学习、深度学习等技术，构建了从数据采集预处理到模型评估优化的完整体系。该模型有效提升了故障预测的精准度与时效性，为应对海上油田复杂环境下的设备故障问题提供了技术支撑。研究成果对于降低生产风险、提升运营效率具有显著贡献，并进一步促进了海上油气开发向智能化的转型进程。

参考文献

- [1] 邓俊武, 张玉民, 张红娣, 杜晓坤. X尾翼无人机的故障诊断和容错控制方法[J]. 山东大学学报(工学版), 2021, 47(05):166-172.
- [2] 冯国强, 赵晓林, 高关根, 寇磊. 无人机系统自平衡容错控制与故障诊断[J]. 计算机测量与控制, 2022, 26(07):21-25.
- [3] 张蒙. 海上油田油井电潜泵故障分析与管理提升[J]. 设备管理与维修, 2023(22):3-4.